**BOOT CAMP 1**

**Data Pipeline for Customer Account Analysis**

This document outlines a data pipeline for processing customer account data, including copying data from a backend team's storage account, performing transformations in Databricks, and upsert data (insert or update) from a file(GOLD Storage) in Azure Data Lake Storage (ADLS) into a dedicated SQL pool table in Azure Synapse Analytics.

**Architecture:**

**A diagram of a factory

AI-generated content may be incorrect.**

**Role Title: Data Engineer (ETL & Transformations on Databricks)**

**Role Summary:** As a Data Engineer, I am responsible for developing and maintaining efficient ETL pipelines and data transformations within Databricks notebooks. My primary focus is on processing large datasets, applying complex data transformations, and ensuring data integrity. I leverage Databricks’ capabilities to optimize workflows, enabling seamless data integration across various sources. Additionally, I manage secure access by providing necessary access tokens, ensuring that team members and stakeholders have the appropriate permissions to interact with the data. I play a crucial role in building a robust data processing framework for the organization.

Backend data with all the required data has been created

A screenshot of a computer

AI-generated content may be incorrect.

Using Get metadata and for each activities, copying the backend data to raw-bronze layer

A screenshot of a computer

AI-generated content may be incorrect.

Files has been successfully reflecting in raw-bronze layer

A screenshot of a computer

AI-generated content may be incorrect.

App created for service principle, application ID and Directory ID

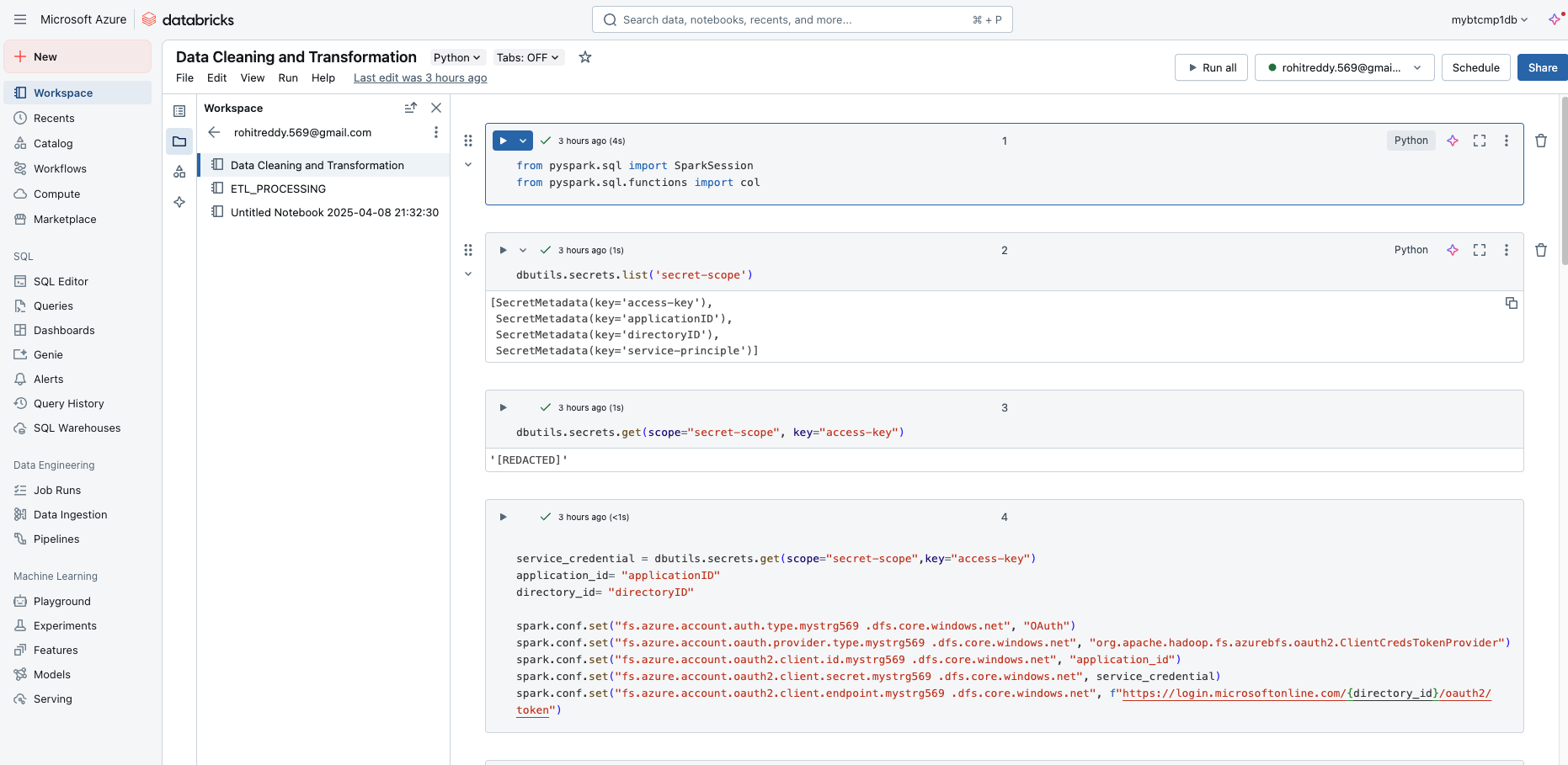
A screenshot of a computer

AI-generated content may be incorrect.

Secret scope has been created with data bricks instance

A screenshot of a computer

AI-generated content may be incorrect.

Data bricks instance has been created with the secret keys   
  
Data cleaning and transformation in Data bricks   
A screenshot of a computer

AI-generated content may be incorrect.

Writing the cleaned data to Silver layer   
A screenshot of a computer

AI-generated content may be incorrect.

Files has been created after the pipeline run in silver container   
A screenshot of a computer

AI-generated content may be incorrect.

ETL Processing in Data bricks   
A screenshot of a computer

AI-generated content may be incorrect.

Got the curated data and writing it to gold layer   
A screenshot of a computer

AI-generated content may be incorrect.

New file has been created in parquet format after the pipeline run  
A screenshot of a computer

AI-generated content may be incorrect.

External tables have been created in synapse analytics   
A screenshot of a computer

AI-generated content may be incorrect.  
Pipeline in ADF has been successfully completed and data can be seen in synapse analytics   
A screenshot of a computer

AI-generated content may be incorrect.

If total balance of all customer id’s are required

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

If balances of specific Customer Id is required

A screenshot of a computer

AI-generated content may be incorrect.